Assignment-Regression Algorithm

1) **Identify your problem statement**
   - Requirements are clear
   - Well defined input column
   - Expect prediction label defined and numerical

     Stage 1 : Machine Learning
     Stage 2 : Supervised Learning
     Stage 3 : Regression

2) **Tell basic info about the dataset (Total number of rows, columns)**
   Input File Name : insurance_pre.csv
   DataFrame dimensions (rows, columns): (1338, 6)

3) **pre-processing method:**
   The variables 'sex' and 'smoker' are nominal and should be converted using one-hot encoding for serialization.

4) **Multiple Linear Regression**
   Best R_Value is 0.7894429387120753

5) **Support Vector Regression – Best R_value (0.760731)**

kernal	C	Gamma	R_VALUE
linear	10	scale	0.001365
linear	100	scale	0.543693
linear	1000	scale	0.634845
linear	2000	scale	0.696461
linear	3000	scale	0.760731
rbf	10	scale	-0.080751
rbf	100	scale	-0.12301
rbf	1000	scale	-0.113902
rbf	2000	scale	-0.100703
rbf	3000	scale	-0.088663
sigmoid	10	scale	-0.090667
sigmoid	100	scale	-0.11762
sigmoid	1000	scale	-2.00648
sigmoid	2000	scale	-7.273098
sigmoid	3000	scale	-15.576208
poly	10	scale	-0.094447
poly	100	scale	-0.09608
poly	1000	scale	-0.019837
poly	2000	scale	0.062006
Poly	3000	scale	0.139372

6) Decision Tree – Best R_Value (0.779097)

CRITERION	MAX_FEATURES	N_ESTIMATORS	R_VALUE
squared_error	None	best	0.708378
squared_error	None	random	0.678705
squared_error	sqrt	best	0.694639
squared_error	sqrt	random	0.642604
squared_error	log2	best	0.693956
squared_error	log2	random	0.710177
friedman_mse	None	best	0.694004
friedman_mse	None	random	0.647651
friedman_mse	sqrt	best	0.689322
friedman_mse	sqrt	random	0.684265
friedman_mse	log2	best	0.673284
friedman_mse	log2	random	0.645677
absolute_error	None	best	0.657153
absolute_error	None	random	0.713341
absolute_error	sqrt	best	0.679758
absolute_error	sqrt	random	0.779097
absolute_error	log2	best	0.67863
absolute_error	log2	random	0.676687
poisson	None	best	0.715657
poisson	None	random	0.680329
poisson	sqrt	best	0.680556
poisson	sqrt	random	0.691287
poisson	log2	best	0.733695
poisson	log2	random	0.651398

7) Random Forest Regressor – Best R_Value (0.865456)

CRITERION	MAX_FEATURES	N_ESTIMATORS	R_VALUE
squared_error	None	random	0.84936
squared_error	sqrt	random	0.86108
squared_error	log2	random	0.865456
friedman_mse	None	random	0.843755
friedman_mse	sqrt	random	0.861197
friedman_mse	log2	random	0.86362
absolute_error	None	random	0.844536
absolute_error	sqrt	random	0.862427
absolute_error	log2	random	0.861128
poisson	None	random	0.851383
poisson	sqrt	random	0.862482
poisson	log2	random	0.862232

Random Forest Regressor yields the highest R value among all models for this input.